Fluid approximation based analysis for mode-switching population dynamics

Citation for published version:
Piho, P & Hillston, J 2021, 'Fluid approximation based analysis for mode-switching population dynamics', *ACM Transactions on Modeling and Computer Simulation*, vol. 31, no. 2, 8. https://doi.org/10.1145/3441680

Digital Object Identifier (DOI):
10.1145/3441680

Published In:
ACM Transactions on Modeling and Computer Simulation

Fluid approximation based analysis for mode-switching population dynamics

PAUL PIHO, University of Edinburgh
JANE HILLSTON, University of Edinburgh

Fluid approximation results provide powerful methods for scalable analysis of models of population dynamics with large numbers of discrete states and have seen wide ranging applications in modelling biological and computer-based systems and model checking. However the applicability of these methods relies on assumptions that are not easily met in a number of modelling scenarios. This paper focuses on one particular class of scenarios in which rapid information propagation in the system is considered. In particular, we study the case where changes in population dynamics are induced by information about the environment being communicated between components of the population via broadcast communication. We see how existing hybrid fluid limit results, resulting in piecewise deterministic Markov processes, can be adapted to such models. Finally, we propose heuristic constructions for extracting the mean behaviour from the resulting approximations without the need to simulate individual trajectories.

CCS Concepts: • **Mathematics of computing → Markov processes**; • **Computing methodologies** → *Agent / discrete models*; • **Theory of computation** → *Random walks and Markov chains*.

Additional Key Words and Phrases: fluid approximation, stochastic modelling, population dynamics, hybrid models

ACM Reference Format:
Paul Piho and Jane Hillston. 2020. Fluid approximation based analysis for mode-switching population dynamics. 1, 1 (November 2020), 29 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Stochastic process models are often used to study real-life complex systems where the dynamics of the system result from interactions between large numbers of individual components. Examples range from models in systems biology to computer networks and queueing. In particular, continuous time Markov chains (CTMCs) are a common target mathematical model supported by many high-level formal specification languages like Chemical Reaction Networks [17, 20], Petri nets [2, 27] and process algebras [4, 23], that aim to simplify the creation of models.

Authors' addresses: Paul Piho, paul.piho@ed.ac.uk, University of Edinburgh, Edinburgh, UK; Jane Hillston, jane.hillston@ed.ac.uk, University of Edinburgh, Edinburgh, UK.

XXXX-XXXX/2020/11-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Most of the CTMC-based modelling frameworks mentioned offer natural ways to describe behaviours of populations by specifying the behaviours of individual components within the population. However, like all discrete state modelling paradigms, CTMCs suffer from the state space explosion problem making the exact methods, based on solutions to Chapman-Kolmogorov equations, prohibitively expensive in practice. Without any extra information about the structure of the CTMC a common approach to overcome this problem is to resort to sampling methods (Gillespie's stochastic simulation algorithm [17]) to gather Monte Carlo trajectories of the model and then calculate summary statistics. For complex models this is often still computationally very expensive. In cases where we have extra information about the model structure we can attempt to study the models via model approximation methods, where the original model is replaced with a simpler alternative that aims to give a faithful approximation of the measures and characteristics of interest. A large body of work exists connected to a special case of CTMCs where the structure of the model corresponds to that of interacting populations (population CTMCs, pCTMCs). Such populations are assumed to consist of a large collection of interacting components, or agents, that are distinguishable only through their state, with the state space of the underlying pCTMC giving the counts of agents in each of the possible states. For a recent review starting from the Chemical Reaction Network viewpoint of population CTMCs please refer to [32]. In particular, for population CTMCs the methods of fluid and central limit approximation [16] offer a way to approximate the dynamics of the population CTMC with a set of ordinary differential equations (ODEs). These techniques have successfully been applied to model checking [5, 8, 9] as well as control synthesis [1, 14] resulting in analysis scalable to systems with large numbers of interacting components. The papers [7, 26] deal with extensions of fluid approximation results to situations where the standard theoretical assumptions do not apply. In the case of [7] this involves dealing with hybrid behaviour where not all components of the system are present in large quantities, while [26] considers general CTMCs without a population structure.

However, the issue of scalability remains relevant, for example, when modelling the effects of communication and information available to the agents within a population on the macro-level behaviour of the population. In particular, consider situations where the agents are equipped with knowledge; moreover assume that they are learning about their environment through experience, and sharing that information within the population through broadcast communication. We assume that the behaviour of the agents may change as they gain in knowledge. A consequence of this is that an action of a single agent (a broadcast) can change the macro-level dynamics of the whole population. For example, the objectives of agents may change as more information is acquired, leading to a change in the overall behaviour of the population. In this paper we focus our attention on cases of models which incorporate such information cascades and consequent shifts in behaviour.

In previous our work [30], such changes in the macro-level behaviour were identified as *mode-switching population dynamics*, a particular form of hybrid model. In this case, levels of knowledge and the related dynamics correspond to the different dynamic modes of the population system. The aim in this paper is to provide a rigorous treatment of such mode-switching population dynamics from the fluid approximation point of view, making them amenable to efficient ODE-based analysis techniques. A wealth of motivating examples can be drawn from the area of mean-field methods for control of swarm robotics [15] where the described information propagation dynamics have not yet been extensively studied.

The contributions of this paper are the following. We develop a formal framework in which to study approximation methods for models with mode-switching population dynamics in the context of existing hybrid fluid approximation results [7, 12]. This is challenging because standard broadcast communication between components, and the subsequent mode switches do not easily fit into the established fluid approximation based methods, and there are subtle differences between these models and those previously considered when fluid approximations of hybrid systems have been established. Secondly, we examine the computational challenges of applying the existing theoretical results and develop pragmatic approximations to overcome these challenges. In particular, we see how an approximation for the mean dynamics of the hybrid fluid approximation can be constructed iteratively, avoiding explicit simulation of the hybrid model.

This paper is structured as follows. We start by presenting the technical background in Section 2. Section 3 introduces the class of models under study – namely mode-switching population systems. Section 4 presents the relevant hybrid fluid approximation results and discusses their applicability to mode-switching populations. Section 5 considers the constructions of marginal mean dynamics corresponding to the population measures. In Section 6 we present the empirical analysis of the constructed approximations for three examples inspired by robot swarms. Finally, we provide related work and conclusions in Section 7 and Section 8 respectively.

2 BACKGROUND

In this section we briefly set up the definitions and some technical background from stochastic modelling. The discussion in this paper is focussed on continuous time Markov processes and in particular CTMCs. The definitions that follow are standard and, for example [28], can be used as a more detailed reference.

A Markov process is a time-indexed family of random variables $\{X(t), t \in [0, \infty)\}$ such that the future behaviour of the process is not dependent on its past. In the rest of the paper we are going to employ the notation X to denote a Markov process. The process X is a Markov process if for all collections of times $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$ and states $i_1, i_2, \ldots, i_n$ we have

$$\mathbb{P}(X(t_n) = i_n \mid X(t_{n-1}) = i_{n-1}, \ldots X(t_0) = i_0) = \mathbb{P}(X(t_n) = i_n \mid X(t_{n-1}) = i_{n-1})$$

A particular example of a Markov process that is going to be the topic throughout this paper is the continuous time Markov chain (CTMC). A CTMC is defined via its infinitesimal generator matrix describing rates at which the CTMC moves between its states. For any $t \in [0, \infty)$ let Q be a matrix with (i, j)-th entry defined by $q(i, j)$ such that the following properties hold.

(1) $0 \leq -q(i, i) < \infty$
(2) $0 \leq q(i, j)$ for $i \neq j$
(3) $\sum_j q(i, j) = 0$

With that in mind we can define a CTMC in terms of its Q matrix in the following way [28].

Definition 1 (Continuous time Markov chain). *Let X be a Markov process with values in a countable set E. Write $Q = (q(i,j) \mid i,j \in E)$ for the associated generator matrix. For all $n \geq 0$, all times $0 \leq t_0 \leq t_1 \leq \cdots \leq t_{n+1}$ and all states $i_0, \cdots, i_{n+1}$*

$$\mathbb{P}(X(t_{n+1}) = i_{n+1} \mid X(t_0) = i_0, \cdots, X(t_n) = i_n) = p(i_{n+1}; t_{n+1} \mid i_n; t_n)$$

satisfying the following Kolmogorov forward equation

$$\frac{\partial}{\partial t} p(j;t \mid i;s) = \sum_k p(k;t \mid i;s) q(k,j), \qquad \text{on } (s,\infty) \text{ with } p(j;s \mid i;s) = \delta_{ij} \tag{1}$$

where δ_{ij} is the Kronecker delta taking the value 1 if i and j are equal and taking the value 0 otherwise. Equation 1 is often referred to in biochemical modelling literature as the chemical master equation.

When the generator matrices are constant and not dependent on time, the CTMC is said to be *time-homogeneous*.

Example 1. *We can consider the Poisson process, which is usually interpreted as a counting process for arrival events, as the simplest example of a time-homogeneous CTMC. For that suppose the CTMC X takes values in the set $\{0,1,2,\ldots\}$ with the generator matrix Q defined by*

$$q(i,j) = \begin{cases} \lambda & \text{for } j = i+1 \\ -\lambda & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The defined process corresponds exactly to the Poisson process with arrival rate λ.

2.1 Population processes

A particular class of time-homogeneous CTMCs is defined through the following formalisation of population dynamics. Consider a system of components evolving in a finite state space $S = \{1, \ldots, K\}$ with components only distinguishable through their state. With that in mind a corresponding population CTMC (pCTMC) can be defined by a tuple $\mathcal{P} = (\boldsymbol{X}, \mathcal{T}, \boldsymbol{X}_0)$ where

- $\boldsymbol{X} = (X_1, \ldots, X_K) \in \mathbb{Z}^K_{\geq 0}$ is a variable with the i-th entry representing the current count of components in state $i \in S$.
- $\mathcal{T}$ is a set of transitions of the form $\tau = (r_\tau, \boldsymbol{v}_\tau)$ where
  - $r_\tau : \mathbb{Z}^K_{\geq 0} \to \mathbb{R}_{\geq 0}$ is the rate function associating transition τ with the rate of an exponential distribution depending on the state of the model.
  - $\boldsymbol{v}_\tau : \mathbb{Z}^K_{\geq 0} \to \mathbb{Z}^K$ is an update which gives the net change for each population variable in $\boldsymbol{X}$ caused by transition τ. In most common uses of pCTMCs the update is a constant function that does not depend on the state of the population before the transition. We are going to denote an update corresponding to τ as $\boldsymbol{v}_\tau$ unless the dependence on the population state has to be made explicit.
- $\boldsymbol{X}_0$ is the initial state of the model.

For ease of presentation we are going to consider a special case of conservative population CTMCs where the total number of components in the system remains constant in time. That is, if the total population is N then for all times $t \geq 0$,

$$\sum_{i=1}^{K} X_i(t) = N.$$

We are going to use the notation $\mathcal{P}^N = (\boldsymbol{X}^N, \mathcal{T}^N, \boldsymbol{X}_0^N)$ for such conservative pCTMCs to make the population size explicit. While the assumption that the total number of components remains fixed throughout the evolution of the system might seem strong at first, most instances of engineered collectives can easily be considered under this assumption. From the theoretical perspective this is one of the ways to ensure appropriate scaling of population variables for the fluid approximation result presented in the next section. The state variables of the scaled model are in our case going to correspond to proportions of components in any given state. However the assumption of the population being conservative is by no means the only way to consider fluid approximations and is chosen here because it fits with the theme of broadcast communication in man-made systems and simplifies the presentation of the results. In the context of CRNs one often considers models where the volume in which reactions take place remains fixed through the evolulution of the system. The fluid approximation result that follow can then instead be derived with respect to a scaling where the state space of the scaled system corresponds to concentrations in unit volume [22].

Example 2. *As our running example for the rest of the paper consider a swarm foraging-inspired scenario where robots are looking for a designated target area and gather to it. Let us suppose that the exploration stage of the robots can be modelled by a random walk on a graph. Thus, for a specific example let us consider Figure 1 showing the state transition diagram of individual robots, where edges labelled by the transition rates.*

The pCTMC resulting from considering a population of robots each behaving according to Figure 1 is defined over four variables — one variable for the population counts in each of the four locations $(0, 1), (0, 0), (1, 0)$ *and* $(1, 1)$. *The set of transitions for this pCTMC would consist of updates and associated rates that correspond to a single robot component changing its location. For example, the transition corresponding to a robot moving from* $(0, 0)$ *to* $(1, 0)$ *would have an update vector that decreases the population at location* $(0, 0)$ *and increases the population at location* $(1, 0)$ *by one. The associated rate, according to Figure 1, is given by* $\frac{1}{2} r_m$ *for some fixed parameter* $r_m > 0$.

The approximation results considered in this paper rely on studying the behaviour of conservative pCTMCs for increasing values of population size N. In order to make a meaningful comparison across the values of N we are going to rescale the model $\mathcal{P}^N = (\boldsymbol{X}^N, \mathcal{T}^N, \boldsymbol{X}_0^N)$ to $\hat{\mathcal{P}}^N = (\hat{\boldsymbol{X}}^N, \hat{\mathcal{T}}^N, \hat{\boldsymbol{X}}_0^N)$ such that $\hat{\boldsymbol{X}}^N$ gives the proportion of components in each of the states in S. In the following we are going to describe how such a scaled model is constructed and how this construction is applied to the running example.

Firstly, the rescaled state of the system is given by $\hat{\boldsymbol{X}}^N \stackrel{\text{def}}{=} \frac{1}{N}\boldsymbol{X}^N$. Secondly, the transitions in $\hat{\mathcal{T}}^N$ are of the form $\hat{\tau} = (\hat{r}^N_\tau, \frac{1}{N}\boldsymbol{v}_\tau)$ where $\hat{r}^N_\tau$ is the rate function over scaled state variables. For this construction we only consider population processes where the rates are density dependent. Formally, this means that for all $\tau \in \mathcal{T}^N$ there exists a Lipschitz continuous function $f^N_\tau : \mathbb{R}^K \longrightarrow \mathbb{R}_{\geq 0}$ such that

$$r_\tau(\boldsymbol{X}^N) = N f^N_\tau\left(\frac{\boldsymbol{X}^N}{N}\right)$$

The scaled rate function $\hat{r}^N$ is then given by

$$\hat{r}^N\left(\frac{\boldsymbol{X}^N}{N}\right) = f^N_\tau\left(\frac{\boldsymbol{X}^N}{N}\right)$$

Finally, the constructed scaled pCTMC $\hat{\mathcal{P}}^N$ gives a $\mathbb{R}^K$-valued CTMC $\hat{\boldsymbol{X}}^N$ with the jump intensities given by

$$q^N(\boldsymbol{x}, \boldsymbol{x}') = \sum_{\tau \in \hat{\mathcal{T}}^N \mid \boldsymbol{v}_\tau = \boldsymbol{x}' - \boldsymbol{x}} \hat{r}_\tau(\boldsymbol{x})$$

In particular, the rate of moving from state $\boldsymbol{x}$ to $\boldsymbol{x}'$ is given by adding the rates of transitions which can cause the change of state.

Example 3. *Returning to our running example and supposing we are dealing with a population of N robots let us denote the state of the population CTMC by*

$$\boldsymbol{X}^N = (X^N_{01}, X^N_{00}, X^N_{10}, X^N_{11})$$

where the variable X^N_{ij} denotes the population count at location (i, j). The variables of the corresponding rescaled population CTMC would then be simply given by

$$\hat{\boldsymbol{X}}^N = \frac{1}{N}(X^N_{01}, X^N_{00}, X^N_{10}, X^N_{11})$$

and correspond to the proportion of robots in each of the available locations. The rescaling of update vectors corresponding to the transitions is done analogously by dividing by the size of the population N. Finally, we need to consider the rescaled rates of the transitions. Again, as an example let us consider the transition, denoted τ, corresponding to a robot moving from $(0, 0)$ to $(1, 0)$. The corresponding scaled rate function is then given by the following.

$$r_\tau(\boldsymbol{X}) = \frac{1}{2} r_m X^N_{00} = N\frac{1}{2} r_m \hat{X}^N_{00} \implies \hat{r}^N_\tau(\hat{\boldsymbol{X}}) = \frac{1}{2} r_m \hat{X}^N_{00}$$

2.2 Fluid approximation

As discussed in the introduction, our analysis of the mode-switching dynamics is going to be based on fluid approximation results. In this section we introduce the approximation result, due to Kurtz [21], in the context of conservative pCTMCs. A detailed treatment of the results can be found in [16, 21]. The fluid approximation arises from considering a sequence of continuous time Markov chains $\hat{\boldsymbol{X}}^N$ derived from a scaled population process $\hat{\mathcal{P}}^N$ as $N \longrightarrow \infty$. Under the assumption that f^N_τ

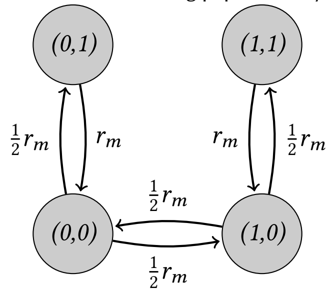


Fig. 1. Single robot – random walk

converges uniformly, as $N \to \infty$, to a locally Lipschitz continuous function f_τ we can construct the limit drift vector

$$F(\hat{\boldsymbol{x}}) = \sum_{\tau \in \hat{\mathcal{T}}^N} \boldsymbol{v}_\tau f_\tau(\hat{\boldsymbol{x}})$$

and state the following theorem.

Theorem 1 (Deterministic approximation theorem [21]). *With $\hat{\boldsymbol{X}}_0^N$ we assume there exists a point $\hat{\boldsymbol{x}}_0$ such that $\lim_{N\to\infty} \|\hat{\boldsymbol{X}}_0^N - \boldsymbol{x}_0\| = 0$ almost surely. Then for every $t \in [0, \infty)$ and $\epsilon > 0$ we have*

$$\mathbb{P}\left(\lim_{N\to\infty} \sup_{u \le t} |\hat{\boldsymbol{X}}^N(u) - \hat{\boldsymbol{x}}(u)| > \epsilon\right) = 0$$

where $\hat{\boldsymbol{x}}$ is a solution to $\frac{d\hat{\boldsymbol{x}}}{dt} = F(\hat{\boldsymbol{x}})$ with $\hat{\boldsymbol{x}}(0) = \hat{\boldsymbol{x}}_0$.

In particular, the discrete stochastic behaviour of the pCTMC is approximated by a continuous deterministic one which corresponds to the limiting behaviour of the stochastic process as the population size N approaches infinity. Observe that the fluid approximation can also be seen as the deterministic drift component of the linear noise approximation.

Example 4. *Applying this to the running example we can write down the limit drift of the rescaled population system modelling a random walk on the graph structure in Figure 1 as follows.*

$$\frac{d}{dt}\hat{\boldsymbol{x}}(t) = \begin{pmatrix} r_m(-\hat{x}_{01}(t) + \frac{1}{2}\hat{x}_{00}(t)) \\ r_m(-\hat{x}_{00}(t) + \frac{1}{2}\hat{x}_{01}(t) + \frac{1}{2}\hat{x}_{10}(t)) \\ r_m(-\hat{x}_{10}(t) + \frac{1}{2}\hat{x}_{11}(t) + \frac{1}{2}\hat{x}_{00}(t)) \\ r_m(-\hat{x}_{11}(t) + \frac{1}{2}\hat{x}_{10}(t)) \end{pmatrix}$$

For a non-random initial condition the above system of ODEs can then be easily solved to get the fluid approximation of the population dynamics. According to Theorem 1 this is the exact limiting behaviour as the population size N tends to infinity. The resulting trajectories and how they compare to stochastic simulation are show in Figure 2.

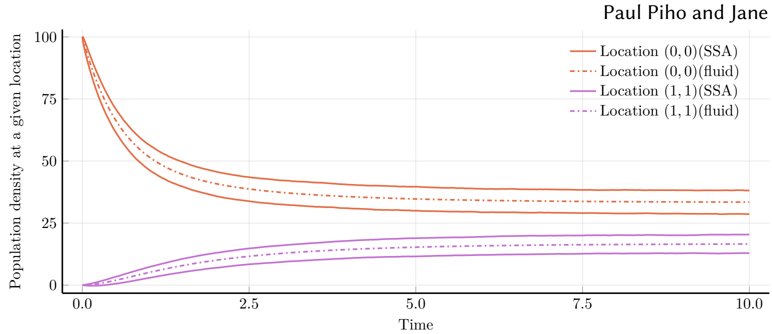


Fig. 2. Fluid approximation trajectories for the population of robots in the running example performing a random walk. The parameter r_m corresponding to the movement rate was set to 1 for this example. For visual comparison, we have given one standard deviation around the empirical mean from 5000 stochastic simulation trajectories.

3 MODE-SWITCHING POPULATION SYSTEMS

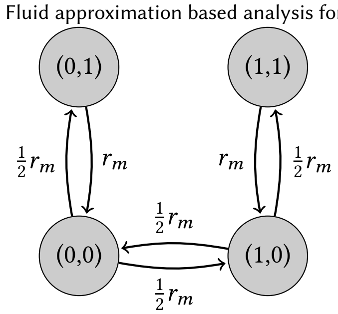
In this section we define a class of stochastic population models for modelling systems where the dynamics of a population can be separated into discrete modes of behaviour. In [30] we argued that such models arise naturally from broadcast communication in collective systems with modes corresponding to information that the collective has about its operating environment. To start off we extend our running example to give the following motivating example.

Example 5. *So far in our running example we have considered the exploration phase of the simple foraging-inspired scenario. We are going to extend this model so that an exploration phase is followed by aggregating to the target area. The switch from the exploration to the aggregation phase happens when an individual in the swarm detects the target area and broadcasts this information to the rest of the swarm. After receiving the information via the broadcast communication each robot in the swarm will perform a directed walk towards the target node. For the example let us consider the target to be location* $(1, 1)$. *The behaviour of the robots in each of the two modes is illustrated in Figure 3. This simple population model can be easily expressed in a high-level process algebraic modelling language resulting in a pCTMC over* 8 *counting variables. The counting variables specify how many robots are present in each of the locations and whether the robots in those locations are following dynamics given in Figure 3a or Figure 3b.*

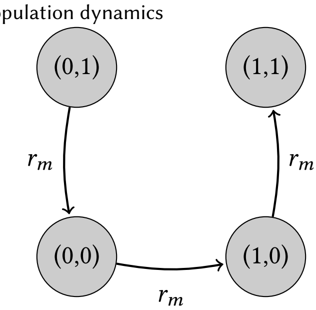
Note that the example described above features two clear dynamic modes. This can be viewed as a particular instance of the class of stochastic models that we are interested in, *mode-switching population systems.*

Definition 2. *We define a mode-switching population system as a joint Markov processes* $\boldsymbol{Y}(t) = (\boldsymbol{X}(t), Z(t)) \in \mathbb{Z}^n_{\geq 0} \times \mathbb{Z}_{\geq 0}$ *where we make the assumption that each transition only changes the state of* $\boldsymbol{X}(t)$ *or* $Z(t)$ *but not both. Moreover, suppose that*

- *conditional on* $Z = z$, *the process* $\{\boldsymbol{X}(t)), t \geq 0\}$ *is given by a population process* $\mathcal{P}_z = (\boldsymbol{X}_z, \mathcal{T}_z, \boldsymbol{X}_0)$ *with jump intensities* $q^z_{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{x}')$.



(a) Before broadcast – random walk



(b) After broadcast – directed walk

Fig. 3. Behaviour of individuals in the swarm model with 4 locations. The parameters r_m and r_s give the movement rate between locations and the sensing rate respectively.

- *conditional on $\boldsymbol{X} = \boldsymbol{x}$, the process $\{Z(t), t \geq 0\}$ is a continuous time Markov chain with intensities $q_Z^{\boldsymbol{x}}(z, z')$.*

The transition intensities of transitioning from state $(\boldsymbol{x}, z)$ to $(\boldsymbol{x}', z')$ for the joint process $\boldsymbol{Y}(t) = (\boldsymbol{X}(t), Z(t))$ are thus defined as

$$q(\boldsymbol{x}, \boldsymbol{x}', z, z') = \begin{cases} q_X^z(\boldsymbol{x}, \boldsymbol{x}') & \text{for } z = z' \\ q_Z^{\boldsymbol{x}}(z, z') & \text{for } \boldsymbol{x} = \boldsymbol{x}' \\ 0 & \text{otherwise} \end{cases}$$

The definition above has similarities to hybrid systems considered in the context of biochemical systems under decomposition of the state-space into low and high-copy number components (see Section 7 for further discussion).

4 LIMITS FOR MODE-SWITCHING POPULATION SYSTEMS

In order to analyse the mode-switching population system introduced in the previous section we are going focus on fluid approximation-based methods motivated by application in mean-field control methods in swarm behaviours [15]. Note that Bortolussi in [7] presents a comprehensive set of results on fluid limits of population CTMCs exhibiting hybrid behaviour where the limiting behaviour is given in terms of piecewise deterministic Markov Processes (PDMP) [13]. For example, guarded behaviour, instantaneous transitions and stochastic jumps coupled with a population structure were considered.

Due to the construction of the mode-switching dynamics the most general treatment of PDMPs is unnecessary for the discussion that follows. An aspect of the constructed models that simplifies the analysis is that the transitions of the mode-switching process do not change the state of the population variables. With this in mind we start by giving a definition of a subclass of PDMPs sufficient here (a similar restriction was also considered in [6]). The intuition is that a PDMP defines a process with continuous deterministic dynamics interrupted by random switching.

Definition 3 ((Simplified) piecewise deterministic Markov process (PDMP) [13]). *Let E be a countable set and $M \subset \mathbb{R}^n$ a compact subset giving the domain of the continuous evolution. We assume that*

- *for all $e \in E$ we have a smooth time-independent vector field $\boldsymbol{F}^e : M \longrightarrow \mathbb{R}^n$ such that the ODE $\frac{d}{dt}\boldsymbol{y}(t) = \boldsymbol{F}^e(\boldsymbol{y}(t))$, for each initial condition $\boldsymbol{y}(0) \in M$, has a unique solution such that $\boldsymbol{y}(t) \in M$ for all $t \geq 0$.*
- *all states $z, z' \in E$ are identified with unit vectors $\boldsymbol{e}_z, \boldsymbol{e}_{z'}$ in $\mathbb{R}^{|E|}$ and for all $\boldsymbol{x} \in M$ there is a continuous time Markov chain defined via the infinitesimal generator matrix $Q^{\boldsymbol{x}}$ with entries $q^{\boldsymbol{x}}(z, z')$.*

We define the corresponding PDMP as a stochastic process $(\boldsymbol{X}(t), Z(t)) \in M \times E$ satisfying the equation

$$\begin{pmatrix} \boldsymbol{X}(t) \\ Z(t) \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}(0) + \int_0^t \boldsymbol{F}(\boldsymbol{X}(s))ds \\ Z(0) + \sum_z \sum_{z' \neq z} (\boldsymbol{e}_{z'} - \boldsymbol{e}_z) N_{zz'} \left(\int_0^t \bar{q}^{X(s)}(Z(t), z, z'))ds \right) \end{pmatrix}$$

where

$$\boldsymbol{F}(X(s)) = \boldsymbol{F}^z(X(s)) \qquad \text{if} \quad Z(s) = \boldsymbol{e}_z$$
$$\bar{q}^{X(s)}(Z(s), z, z') = \begin{cases} q^{X(s)}(z, z') & \text{if} \quad Z(s) = \boldsymbol{e}_z \\ 0 & \text{otherwise} \end{cases}$$

and each $N_{zz'}$ is a time-inhomogeneous Poisson process with rate 1 counting the number of transitions from state z to z'.

We have given the process Z in terms of the Poisson representation of Markov chains [16] and constructed it so that the state of it is always given by a unit vector in $\mathbb{R}^{|E|}$. The time parameter is transformed $t \mapsto \bar{q}^{X(s)}(Z(s), z, z')$ based on the state of X and Z so that the process $N_{zz'}$ counts transitions from z to z' only. This gives us a compact definition of PMDPs. In the rest of the paper we are going to use a shorthand and describe the state space of Z by integer values. However, the underlying construction as a process taking values in $\mathbb{R}^{|E|}$ remains in place. Details and the definition of PDMPs in a more general setting can be found in [13]. For limit results in [7] the more general definition was considered, for example, to deal with instantaneous transitions.

In the following we are going to apply the results of [7] to conservative mode-switching population systems as introduced in Definition 2. We denote the joint process corresponding to a mode-switching population process of a fixed number of components by $(\boldsymbol{X}^N, Z^N)$ to indicate the population size N. For each state z of the mode-switching process Z^N we have a pCTMC $\mathcal{P}_z^N = (\boldsymbol{X}_z^N, \mathcal{T}_z^N, \boldsymbol{X}_0^N)$ corresponding to the population dynamics for the given mode z. We are going to consider the following conditions:

(1) The pCTMCs $\mathcal{P}_z^N$ scale for all z as described in Section 2.1. In particular, for all z consider the scaled pCTMC $\hat{\mathcal{P}}_z^N = (\hat{\boldsymbol{X}}_z^N, \hat{\mathcal{T}}_z^N, \hat{\boldsymbol{X}}_0^N)$.
(2) For the intensities $q_Z^{\boldsymbol{x},N}(z, z')$ of the process Z^N conditional on the state of the population $\boldsymbol{X}^N = \boldsymbol{x}$ we require that there exists a Lipschitz continuous function $f_{zz'}^N : \mathbb{R}^K \longrightarrow \mathbb{R}$ such that

$$q_Z^{\boldsymbol{x},N}(z, z') = f_{zz'}^N \left(\frac{\boldsymbol{x}}{N} \right)$$

and $f_{zz'}^N$ converges uniformly to a continuous function $f_{zz'}$ as $N \longrightarrow \infty$. Thus, the scaling for the mode-switching system or discrete variables follows different rules than the scaling for population variables. In particular, transition rates for the discrete variables vary with the scaled population variables rather than the number of components that are part of a given transition.

If the above conditions are satisfied then we can state the hybrid fluid limit theorem in the context of the mode-switching population system as follows.

Theorem 2 (Hybrid fluid limit [7]). *Let $(\boldsymbol{X}^N, Z^N)$ be a mode switching population system satisfying the scaling conditions above and let $(\hat{\boldsymbol{X}}^N, \hat{Z}^N)$ be the corresponding scaled process. Then the sequence of scaled processes $\{(\hat{\boldsymbol{X}}^N, \hat{Z}^N)\}$ converges weakly to a process $(\hat{\boldsymbol{x}}, \hat{Z})$*

$$\begin{pmatrix}\hat{\boldsymbol{x}}(t)\\ \hat{Z}(t)\end{pmatrix} = \begin{pmatrix}\hat{\boldsymbol{x}}_0 + \int_0^t \boldsymbol{F}(\hat{\boldsymbol{x}}(s))ds \\ \hat{Z}(0) + \sum_z \sum_{z' \neq z} (\boldsymbol{e}_{z'} - \boldsymbol{e}_z) N_{z'z}\left(\int_0^t \bar{f}_{zz'}(\hat{\boldsymbol{x}}(s))ds\right)\end{pmatrix}$$

where

$$\boldsymbol{F}(\hat{\boldsymbol{x}}(s)) = \sum_{\tau \in \hat{\mathcal{T}}_z} f_\tau(\hat{\boldsymbol{x}}(s)) \qquad \textit{if} \quad \hat{Z}(s) = z$$

$$\bar{f}_{zz'}(\hat{Z}(s), \hat{\boldsymbol{x}}(s)) = \begin{cases} f_{zz'}(\hat{\boldsymbol{x}}(s)) & \textit{if} \quad \hat{Z}(s) = z \\ 0 & \textit{otherwise}\end{cases}$$

The rates f_τ for $\tau \in \hat{\mathcal{T}}_z$ are given by the constructions in Section 2.1 applied to the pCTMC $\hat{\mathcal{P}}_z^N = (\hat{\boldsymbol{X}}_z^N, \hat{\mathcal{T}}_z^N, \hat{\boldsymbol{X}}_0^N)$ defining the process $\hat{\boldsymbol{X}}^N$ conditional on the state z of the mode-switching process. The process is an example of a PDMP [13] as in between stochastic jumps of $\hat{Z}$ the state of $\hat{Z}$ is constant while the evolution of $\hat{\boldsymbol{x}}$ is given with respect to a deterministic drift.

Proof. The idea of the proof is to consider the limiting behaviour of the process $\hat{\boldsymbol{X}}^N$ inductively between the stochastic jumps of the process Z^N. The details of the proof will not be recreated here and can be found in [7]. □

For a case like the running example we are not quite ready to apply the above theorem yet. To see that let us consider the following example.

Example 6. *Suppose each of the robots that reaches the location $(1, 1)$ is capable of causing the mode switch at some rate r_s. Let the mode corresponding to exploration be denoted by $Z = 0$ and the mode corresponding to gathering to the target location be denoted by $Z = 1$. The total rate at which the mode switch happens is given by the number of robots at $(1, 1)$, denoted X_{11}, multiplied by the rate r_s. That is, we get*

$$q_Z^{\boldsymbol{X},N}(0, 1) = r_s X_{11}$$

which does not satisfy the scaling requirement defined above for the mode switches due to the dependence on the non-scaled population variables. Considering the limiting behaviour for such transitions would not lead to anything interesting. In particular, taking $N \longrightarrow \infty$, the limiting behaviour of the corresponding mode switching process is such that the probability

$$\mathbb{P}(\textit{broadcast has happened by time } t) \longrightarrow 1 \qquad \textit{for all} \quad t > 0$$

Thus, the limit behaviour of the mode-switching process corresponding to a broadcast message being sent is expected to immediately reach its absorbing state. While the limit of the switching population constructed in such a way is valid it is not very useful when our aim is to understand the behaviour of the system at a fixed finite population size N.

In the following we construct an approximation dependent on the population size. To that end, in order to leverage Theorem 2 in the case of running example we start by constructing a special instance of the mode-switching process Z^N that fixes the behaviour of the mode-switching process to a given population size.

Suppose we are interested in approximate dynamics of the mode-switching population $(\boldsymbol{X}^N, Z^N)$ for a fixed population size $\tilde{N}$. The choice of $\tilde{N}$ is entirely up to the modeller and is chosen based on the modelling problem at hand. For example, if we are interested in the approximate behaviour of the system with 100 homogeneous individual components we would set $\tilde{N} = 100$. In order to construct an approximation for the population dynamics at the chosen population level suppose that for intensities $q_Z^{\boldsymbol{x},N}(z, z')$ of the mode-switching process Z^N given the state of the population process $\boldsymbol{X}^N = \boldsymbol{x}$ there exists a Lipschitz continuous function $f_{zz'}^N : \mathbb{R}^K \to \mathbb{R}$ such that

$$q_Z^{\boldsymbol{x},N}(z, z') = N f_{zz'}^N \left(\frac{\boldsymbol{x}}{N}\right)$$

and $f_{zz'}^N$ converges uniformly to a continuous function $f_{zz'}$ as $N \to \infty$. As noted before this does not satisfy the scaling requirements for the discrete variables that would allow us to make use of the hybrid fluid approximation directly. To overcome this problem we construct a special instance $\hat{Z}^N$ of Z^N with intensities given by

$$q_{\hat{Z}}^{\boldsymbol{x},N}(z, z') = \tilde{N} f_{zz'}^N \left(\frac{\boldsymbol{x}}{N}\right)$$

In particular,

- the density-dependent scaling for rates $f_{zz'}^N$ is done according to the population size N which is later taken to the limit $N \to \infty$.
- multiplied by the chosen population level $\tilde{N}$ which is going to be kept constant.

Effectively this fixes the behaviour of the mode-switching process $\hat{Z}^N$ to the case where, for the purpose of switching, the total population size is assumed to be $\tilde{N}$.

Example 7. *Let us consider the running example. If again the population size at the location* $(1, 1)$ *is* X_{11} *then the broadcast happens with rate* $r_s X_{11}$*. Similarly, the broadcast for the scaled population process given population density* $\frac{X_{11}}{N}$ *at location* $(1, 1)$ *happens with the rate* $N r_s \frac{X_{11}}{N}$*. The special instance of the joint process at population level* 100 *according to the construction above is then defined by saying that the rate of broadcast is given by* $100 r_s \frac{n}{N}$*. This now satisfies the scaling condition for the mode-switching transitions.*

Note that the scaled joint process $(\hat{\boldsymbol{X}}^N, \hat{Z}^N)$, constructed from $(\boldsymbol{X}^N, Z^N)$ via taking the special instance of Z^N, satisfies the conditions of Theorem 2 and thus converges weakly, as $N \to \infty$ to the PMDP $(\hat{\boldsymbol{x}}, \hat{Z})$ with $\hat{Z}$ defined by the intensities $q_{\hat{Z}}^{\boldsymbol{i},N}$ as $N \to \infty$. As mentioned, the motivation here is that while the behaviour of the scaled population process is taken to its asymptotic limit the behaviour of the mode-switching process is kept fixed to correspond to a chosen population level $\tilde{N}$.

Example 8. *For the running example we can easily see that the deterministic behaviour is given by the following drifts.*

$$\frac{d}{dt}\hat{\boldsymbol{x}}(t) = \begin{cases} \begin{pmatrix} r_m(-\hat{x}_{01}(t) + \frac{1}{2}\hat{x}_{00}(t)) \\ r_m(-\hat{x}_{00}(t) + \frac{1}{2}\hat{x}_{01}(t) + \frac{1}{2}\hat{x}_{10}(t)) \\ r_m(-\hat{x}_{10}(t) + \frac{1}{2}\hat{x}_{11}(t) + \frac{1}{2}\hat{x}_{00}(t)) \\ r_m(-\hat{x}_{11}(t) + \frac{1}{2}\hat{x}_{10}(t)) \end{pmatrix} & \text{for } \hat{Z}(t) = 0 \\ \\ \begin{pmatrix} -r_m\hat{x}_{01}(t) \\ r_m(-\hat{x}_{00}(t) + \hat{x}_{01}(t)) \\ r_m(-\hat{x}_{10}(t) + \hat{x}_{00}(t)) \\ r_m(\hat{x}_{10}(t)) \end{pmatrix} & \text{for } \hat{Z}(t) = 1 \end{cases} \tag{3}$$

with

$$\hat{\boldsymbol{x}}(t) = \begin{pmatrix} \hat{x}_{01}(t) & \hat{x}_{00}(t) & \hat{x}_{10}(t) & \hat{x}_{11}(t) \end{pmatrix}^T$$

and initial conditions given by

$$\hat{\boldsymbol{x}}(0) = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix}^T$$
$$\hat{Z}(0) = 0$$

If we continue with the same choice of $\tilde{N}$ as in the previous example then the switching from state 0 *of $\hat{Z}$ to state* 1 *of $\hat{Z}$ happens at rate* $100 r_s \hat{x}_{11}$.

The difficulty of treating the approximation numerically still remains as $\hat{Z}$ is a random process depending on $\hat{\boldsymbol{x}}$ while $\hat{\boldsymbol{x}}$ depends on $\hat{Z}$. Although, it is possible to draw realisations of the time-inhomogeneous Poisson processes describing the evolution of the mode-switching process the computational demand of the problem remains high. This is due to the parameters of the Poisson processes depending continuously on the population variables. In the next section, however, we are going to observe that certain constraints on the structure of the mode-switching processes greatly simplify the numerical calculations.

5 MARGINAL DYNAMICS

We start with the observation that, in general, the dynamics of interest are those of the marginal process describing the population. That is, using the notation $(\hat{\boldsymbol{x}}, \hat{Z})$ to denote the hybrid fluid limit of a mode-switching population system, we are interested in the behaviour of the marginal process $\hat{\boldsymbol{x}}$. The naive approach would be to marginalise out the effects of the switching process $\hat{Z}$.

To start let us consider the marginal process $\hat{\boldsymbol{x}}(t)$ given a sample trajectory $\boldsymbol{x}_{s^-}$ in the time interval $[0, s]$. Moreover, let $\boldsymbol{x}_s$ denote the state of sample trajectory $\boldsymbol{x}_{s^-}$ at time s. Denoting the limit drift vector of the process $\hat{\boldsymbol{x}}(t)$ given the state of the mode-switching process is $\hat{Z}(t) = z$ as $\boldsymbol{F}^z$ and the i-th component of the vector as $\boldsymbol{F}_i^z$

we get

$$\frac{\partial}{\partial t}p(\boldsymbol{x};t \mid \boldsymbol{x}_{s^-};s) = \frac{\partial}{\partial t}\sum_{z} p(z;s \mid \boldsymbol{x}_{s^-};s)\sum_{z'} p(\boldsymbol{x},z';t \mid \boldsymbol{x}_s,z;s) \tag{4}$$

$$= \sum_{z} p(z;s \mid \boldsymbol{x}_{s^-};s)\sum_{z'}\Big[-\sum_{i}\partial_i \boldsymbol{F}_i^z(\boldsymbol{x})p(\boldsymbol{x},z';t \mid \boldsymbol{x}_s,z;s) \tag{5}$$

$$+\sum_{k,k\neq z'}\Big[p(\boldsymbol{x},k;t \mid \boldsymbol{x}_s,z;s)q_{\hat{Z}}^{\boldsymbol{x}}(k,z') - p(\boldsymbol{x},z';t \mid \boldsymbol{x}_s,z;s)q_{\hat{Z}}^{\boldsymbol{x}}(z',k)\Big]\Big] \tag{6}$$

where $p(\boldsymbol{x},z';t \mid \boldsymbol{x}_s,z;s)$ denotes the probability that the joint process $(\hat{\boldsymbol{x}},\hat{Z})$ takes the value $(\boldsymbol{x},z')$ at time t and the value $(\boldsymbol{x}_s,z)$ at time $s \leq t$. The above makes use of the characterisation of the forward equation for stochastic hybrid systems [3]

$$\frac{\partial}{\partial t}p(\boldsymbol{x},z';t \mid \boldsymbol{x}_s,z;s) = -\sum_{i}\partial_i \boldsymbol{F}_i^z(\boldsymbol{x})p(\boldsymbol{x},z';t \mid \boldsymbol{x}_s,z;s)$$
$$+\sum_{k,k\neq z'}\Big[p(\boldsymbol{x},k;t \mid \boldsymbol{x}_s,z;s)q_{\hat{Z}}^{\boldsymbol{x}}(k,z') - p(\boldsymbol{x},z';t \mid \boldsymbol{x}_s,z;s)q_{\hat{Z}}^{\boldsymbol{x}}(z',k)\Big]$$

describing the time evolution of the probability density function of the joint system $(\hat{\boldsymbol{x}},\hat{Z})$. Note that here the forward equation gives an exact description of the probability density of the PDMP $(\hat{\boldsymbol{x}},\hat{Z})$ rather than an approximation.

The expression in Equation 6 depends on the continuous evolution given by the limit drifts $\boldsymbol{F}^z$ and discrete jumps of the mode-switching process. Note that the discrete jumps do not have a direct effect on the state of the population variables by Definition 2 and thus, evaluated at $t = s$, we get the following time-evolution of the marginal process where the contribution from the discrete jumps vanishes.

$$\frac{d}{dt}p(\boldsymbol{x};t \mid \boldsymbol{x}_{t^-}) = \mathbb{E}_{\hat{Z}\mid \boldsymbol{x}_{t^-}}\Big[-\sum_{i}\partial_i \boldsymbol{F}_i^{\hat{Z}}(\boldsymbol{x})\Big]$$

The problem here lies in the future behaviour of the marginal process $\hat{\boldsymbol{x}}$ at time t is given in terms of history of the process up to time t. Thus, in order to understand the marginal process $\hat{\boldsymbol{x}}$ we have to reconstruct the stochastic process $\hat{Z}$ at time t from a trajectory $\boldsymbol{x}_{t^-}$. Using the notation

$$\pi_t(z') = p(z';t \mid \boldsymbol{x}_{t^-})$$

to denote the distribution corresponding to the probability of $\hat{Z} = z'$ given the observed history $\boldsymbol{x}_{t^-}$ of the population process $\hat{\boldsymbol{x}}$ can be derived to be

$$\frac{d}{dt}\pi_t(z') = \sum_{z}\pi_t(z)q_Z^{\boldsymbol{x}}(z,z') - \pi_t(z')\Big[\sum_{i}\partial_i \boldsymbol{F}_i^{z'}(\boldsymbol{x}) - \mathbb{E}_{\hat{Z}\mid \boldsymbol{x}_{t^-}}\Big[\sum_{i}\partial_i \boldsymbol{F}_i^{\hat{Z}}(\boldsymbol{x})\Big]\Big]$$

where the first part of the equation can be recognised as the Kolmogorov forward equation for the marginal process $\hat{Z}$. The second part of the equation contains the information gained from observing the trajectory $\boldsymbol{x}_{t^-}$. The contributions of this part are more significant for the states z' that are not too unlikely given the trajectory $\boldsymbol{x}_{t^-}$ (that is, $\pi_t(z')$ not too small) and for which the flow out of state $\boldsymbol{x}$ differs from the expected flow out of $\boldsymbol{x}$ with respect to the filtering distribution. That is, we gain information about the state of $\hat{Z}$ from observing the trajectory of $\hat{\boldsymbol{x}}$ due to the fact that the dynamics of $\hat{\boldsymbol{x}}$ depend on $\hat{Z}$.

The derivation of the filtering distribution follows analogously to the one given for CTMCs in [10]. The full derivation is given for completeness in Appendix A. We are going to follow the established terminology and call the distribution $\pi_t(z')$ a filtering distribution [29].

Example 9. *If the dynamics of the process $\hat{\boldsymbol{x}}$ do not change depending on the state of the process $\hat{Z}$ then part of the equation corresponding to observing the history of the process $\boldsymbol{x}_{t^-}$ becomes* 0*. In particular,*

$$\sum_i \partial_i F_i^{z'}(\boldsymbol{x}) = \mathbb{E}_{\hat{Z}|\boldsymbol{x}_{t^-}}\Big[\sum_i \partial_i F_i^{\hat{Z}}(\boldsymbol{x})\Big]$$

Based on that the construction of a useful approximate description of the marginal process $\hat{\boldsymbol{x}}$ relies on the problem of finding a sufficiently simple description of the filtering distribution. Here we are going to propose the following, as the first order heuristic for deriving approximate dynamics:

$$\frac{d}{dt}\pi_t(z') \approx \sum_z \pi_t(z) q_{\hat{Z}}^{\boldsymbol{x}}(z, z') \tag{7}$$

$$\frac{d}{dt}p(\boldsymbol{x};t) \approx \mathbb{E}_{\hat{Z}}\Big[-\sum_i \partial_i F_i(\boldsymbol{x})\Big] \tag{8}$$

Recall here that the only source of stochasticity that was left in the PDMP approximation of the mode-switching populations resulted from the jump process describing the mode-switching. The result of discarding the information about the history of the process is that we lose all information about the stochasticity of the marginal process. In particular, Equation 8 describes a deterministic system with the time-derivative given by

$$\frac{d}{dt}\hat{\boldsymbol{x}}(t) = \mathbb{E}_{\hat{Z}}\left[\boldsymbol{F}(\hat{\boldsymbol{x}}(t))\right] \tag{9}$$

We are going to use the above as an approximation to the marginal process $\hat{\boldsymbol{x}}$ of the limit PDMP $(\hat{\boldsymbol{x}}, \hat{Z})$. That is, the approximation to the mean of the population marginal process now follows the expectation with respect to the marginal mode-switching process. In the following we are going to consider the computational treatment of this model.

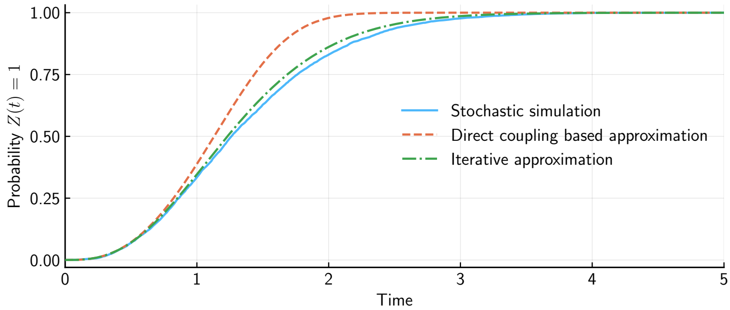
5.1 Direct coupling

The first approach we are going to consider for solving the system given in Equation 9 is to directly couple the approximate filtering distribution

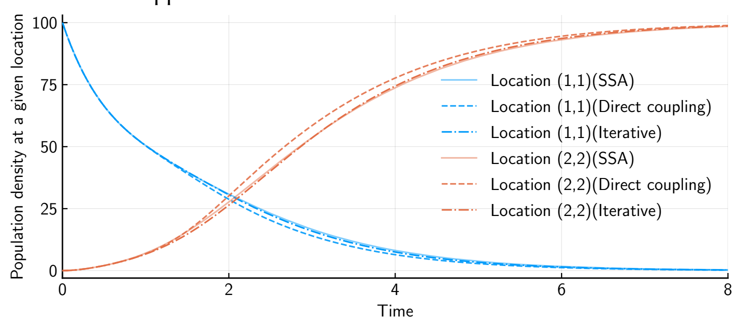
$$\frac{d}{dt}\pi_t(z') = \sum_z \pi_t(z) q_{\hat{Z}}^{\boldsymbol{x}}(z, z')$$

with the approximation of the deterministic drift given by

$$\frac{d}{dt}\hat{\boldsymbol{x}}(t) = \mathbb{E}_{\hat{Z}}\left[\boldsymbol{F}(\hat{\boldsymbol{x}}(t))\right] = \sum_z \pi_t(z) \boldsymbol{F}^z(\hat{\boldsymbol{x}}(t))$$



(a) Comparison of the empirical distribution for $Z = 1$ and the distribution from the fluid approximation solution.



(b) Relative error between the empirical mean for the locations $(0, 0)$ and $(1, 1)$ from 5000 stochastic simulation runs and the approximation constructed from the hybrid fluid approximation.

Fig. 4. Fluid approximation directly coupled with the probability density for mode-switching.

This is done by setting $\frac{d}{dt}\pi_t(z') = \sum_z \pi_t(z) q_{\hat{Z}}^{\hat{x}(t)}(z, z')$. If the number of modes is not too large this can easily be solved as a system of ODEs.

Example 10. *The running example features a single mode switch and thus we can construct the following system of ODEs*

$$\frac{d}{dt}\hat{x}(t) = \pi_t(0)\begin{pmatrix} r_m(-\hat{x}_{01}(t) + \frac{1}{2}\hat{x}_{00}(t)) \\ r_m(-\hat{x}_{00}(t) + \frac{1}{2}\hat{x}_{01}(t) + \frac{1}{2}\hat{x}_{10}(t)) \\ r_m(-\hat{x}_{10}(t) + \frac{1}{2}\hat{x}_{11}(t) + \frac{1}{2}\hat{x}_{00}(t)) \\ r_m(-\hat{x}_{11}(t) + \frac{1}{2}\hat{x}_{10}(t)) \end{pmatrix} + \pi_t(1)\begin{pmatrix} -r_m\hat{x}_{01}(t) \\ r_m(-\hat{x}_{00}(t) + \hat{x}_{01}(t)) \\ r_m(-\hat{x}_{10}(t) + \hat{x}_{00}(t)) \\ r_m(\hat{x}_{10}(t)) \end{pmatrix}$$

$$\frac{d}{dt}\pi_t(0) = -\pi_t(0) r_s \hat{x}_{11}(t)\tilde{N}$$

$$\frac{d}{dt}\pi_t(1) = \pi_t(0) r_s \hat{x}_{11}(t)\tilde{N}$$

where $\tilde{N}$ is the chosen population size. According to the constructions presented in Section 4 the jump rates of the mode-switching process depend on the population size. The construction of population dynamics uses the scaling given for population transitions presented in Section 2.1 and illustrated in Example 3. We have linked the behaviour of the

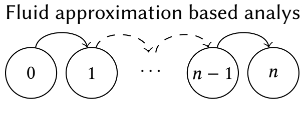


Fig. 5. Pure birth process.

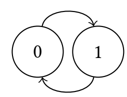


Fig. 6. Stochastic switch.

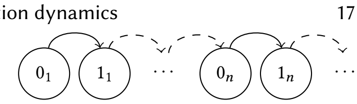


Fig. 7. Pure birth process from stochastic switch.

mode-switching process to a population level of interest $\tilde{N}$. In the following calculations we set $\tilde{N} = 100$. The above system can then be solved using standard ODE solvers for initial conditions

$$\boldsymbol{x}(0) = (0, 1.0, 0, 0) \qquad \pi_t(0) = 1 \qquad \pi_t(1) = 0$$

and parameters $r_m = 1, r_s = 0.2$. Figure 4 gives a visual comparison of the resulting solution with the mean from 5000 runs of the stochastic simulation.

5.2 Iterative method

The method in the previous section of directly coupling the equations for approximate probability density for the marginal process $\hat{Z}$ arising from the heuristic approximations of the filtering equation gives a reasonably good estimate for the mean of the marginal population process in the case of the running example. However, from Figure 4a we can see that even for this simple example the method does not accurately capture the mode-switching dynamics. In this section we present a slightly modified method based on the hybrid fluid approximation that in the case of a certain restricted class of models allow us to more accurately capture the mean dynamics. In particular, we consider mode-switching processes which do not branch.

Non-branching mode-switching processes. In this section, we consider non-branching mode-switching processes $\hat{Z}$, where from each state there is at most one transition out. This will allow us to set up an iterative construction presented in this section. Two examples of such processes are pure birth (or death) processes depicted in Figure 5 and a stochastic switch given in Figure 6. As long as there is no branching the mode-switching processes featuring loops, like the stochastic switch given in Figure 6, can be treated equivalently to pure birth processes. In particular, we can unroll the loops by considering the process describing how many times a given state has been visited. Supposing the stochastic switch is initially in state 0 we would then consider the process depicted in Figure 7

The key characteristic of such non-branching processes is that we can characterise their behaviour in terms of a sequence of first hitting time problems. To see that, let us give a standard definition of first hitting times.

Definition 4. *Let φ be a continuously differentiable function on the space $\mathbb{R}^{K+1}$ (state space of the process $(\boldsymbol{X}, Z)$) with $\varphi(\boldsymbol{X}, Z) > 0$. Let h denote the first hitting time given by*

$$h = \inf\{t \mid \varphi(\boldsymbol{X}(t), Z(t)) \leq 0\}$$

With that in mind, if h_n denotes the first hitting time corresponding to reaching the state n of the pure birth process in Figure 5 then the probability of the process Z

being in state n at time t can be given by

$$p(z = n; t) = p(h_n \leq t, h_{n+1} > t)$$

This corresponds to the joint probability that the state n has been reached before time t but the state $n + 1$ has not. Similarly, conditioned on $h_{n+1} > t$ we have for the pure birth structure that

$$p(z = n; t \mid h_{n+1} > t) = p(h_n \leq t)$$

In the next section we are going to use these observations to iteratively construct the mean dynamics from the hybrid fluid approximation $(\hat{\boldsymbol{x}}, \hat{Z})$ under a simplification of the filtering distribution that discards the history of the process.

Construction of mean dynamics. We propose an iterative method for constructing the marginal dynamics of $\hat{\boldsymbol{x}}$ in a finite time interval $[0, T)$. For a non-branching mode-switching process $\hat{Z}$ let us consider a sequence of first hitting times $h_1, h_2, \ldots, h_n \ldots$ corresponding to mode-switching times into state $1, 2, \ldots n \ldots$ of $\hat{Z}$. As before we are going to leverage the constructed approximate dynamics given by

$$\frac{d}{dt}\pi_t(z') \approx \sum_z \pi_t(z) q^{\boldsymbol{x}}_{\hat{Z}}(z, z') \tag{10}$$

$$\frac{d}{dt}p(\hat{\boldsymbol{x}}; t) \approx \mathbb{E}_{\hat{Z}}\Big[-\sum_i \partial_i \boldsymbol{F}^{\hat{Z}}_i(\hat{\boldsymbol{x}}) \Big] \tag{11}$$

Recall that the sum inside the expectation operator is given over the components of the limit drift vector $\boldsymbol{F}^{\hat{Z}}$ given the state of $\hat{Z}$. As the structure of the mode-switching process is assumed to be non-branching we can conclude that

$$\frac{d}{dt}\hat{\boldsymbol{x}}(t) \approx \sum_j p(h_j < t, h_{j+1} \geq t) \boldsymbol{F}^j(\hat{\boldsymbol{x}}(t)) \tag{12}$$

Note that the sum above is given over the states of $\hat{Z}$. The iterative construction we propose relies on conditioning the behaviour of the population on a limited set of behaviours of $\hat{Z}$ up to some time t. In that case, we let $\hat{Z}^i_{t^-}$ denote the mode-switching process that takes values up to state i within the time interval $[0, t)$ and consider the probability distribution

$$p(z'; t \mid \hat{Z}^i_{t^-})$$

We can then take the derivative with respect to t and consider the time evolution of the above distribution given the mode-switching process is contained within the singleton set $\{0\}$ defined by a non-random initial condition.

$$\begin{aligned}\frac{d}{dt}p(z'; t \mid \hat{Z}^0_{t^-}) &= \frac{d}{dt}\left[\sum_{\boldsymbol{x}} p(z'; t \mid \boldsymbol{x}, \hat{Z}^0_{t^-}; t) p(\boldsymbol{x}; t \mid \hat{Z}^0_{t^-})\right] \\ &= \sum_{\boldsymbol{x}} \left[p(\boldsymbol{x}; t \mid \hat{Z}^0_{t^-}) \frac{d}{dt} p(z'; t \mid \boldsymbol{x}, \hat{Z}^0_{t^-}; t) + p(z'; t \mid \boldsymbol{x}, \hat{Z}^0_{t^-}; t) \frac{d}{dt} p(\boldsymbol{x}; t \mid \hat{Z}^0_{t^-}) \right]\end{aligned}$$

As we have assumed that the mode-switching process does not leave the singleton set $\{0\}$ the term $p(z'; t \mid \boldsymbol{x}, \hat{Z}^0_{t^-}; t)$ is only non-zero if $z' = 0$. Furthermore, the term $p(z'; t \mid \boldsymbol{x}, \hat{Z}^0_{t^-}; t)$ corresponds to our heuristic simplification of the filtering equation under the additional assumption that the mode-switching process $\hat{Z}$ stays in state 0, denoted $\pi_t(z' \mid \hat{Z}^0_{t^-})$. This gives us

$$\begin{aligned}\frac{d}{dt}p(z'=1;t \mid \hat{Z}^0_{t^-}) &= \sum_{\boldsymbol{x}} p(\boldsymbol{x};t \mid \hat{Z}^0_{t^-}) \sum_{z} \pi_t(z \mid \hat{Z}^0_{t^-}) q^{\boldsymbol{x}}_{\hat{Z}}(z,1) \\ &= \mathbb{E}_{\hat{\boldsymbol{x}}|\hat{Z}^0_{t^-}}\left[\pi_t(0 \mid \hat{Z}^0_{t^-}) q^{\boldsymbol{x}}_{\hat{Z}}(0,1)\right] = \mathbb{E}_{\hat{\boldsymbol{x}}|\hat{Z}^0_{t^-}}\left[q^{\boldsymbol{x}}_{\hat{Z}}(0,1)\right] \qquad (13)\end{aligned}$$

The above gives the initial step in our iterative construction. In order to illustrate it let us briefly consider again the running example.

Example 11. *If the population at location* $(1,1)$, *given the mode-switching process is in state* 0 *at time* t, *is denoted as* $x^0_{11}(t)$ *we get*

$$\mathbb{E}_{\hat{\boldsymbol{x}}|\hat{Z}^0_{t^-}}\left[q^{\hat{\boldsymbol{x}}}_{\hat{Z}}(0,1)\right] = q^{\hat{x}^0_{11}(t)}_{\hat{Z}}(0,1)$$

Conditioning the population dynamics on $\hat{Z}^0_{t^-}$ *we can easily find the population variable* $\hat{x}^0_{11}(t)$ *by solving the following system.*

$$\frac{d}{dt}\hat{\boldsymbol{x}}^0(t) = \boldsymbol{F}^0(\hat{\boldsymbol{x}}^0(t)) = \begin{pmatrix} r_m(-\hat{x}_{01}(t) + \frac{1}{2}\hat{x}_{00}(t)) \\ r_m(-\hat{x}_{00}(t) + \frac{1}{2}\hat{x}_{01}(t) + \frac{1}{2}\hat{x}_{10}(t)) \\ r_m(-\hat{x}_{10}(t) + \frac{1}{2}\hat{x}_{11}(t) + \frac{1}{2}\hat{x}_{00}(t)) \\ r_m(-\hat{x}_{11}(t) + \frac{1}{2}\hat{x}_{10}(t)) \end{pmatrix}$$

Let us consider the same non-random initial conditions and parametrisation as previously.

$$\hat{\boldsymbol{x}}(0) = \begin{pmatrix}0 & 1 & 0 & 0\end{pmatrix}^T \qquad \hat{Z}(0) = 0$$

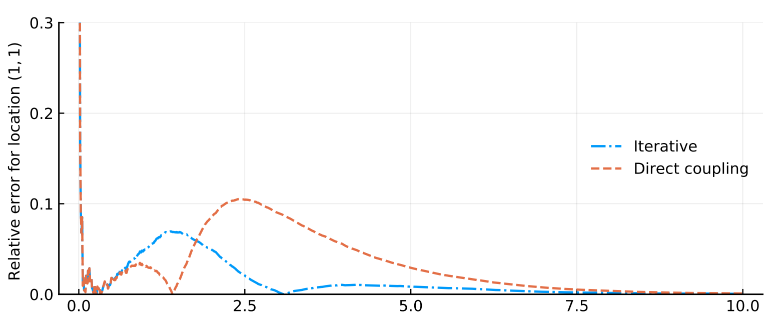


Fig. 8. Relative error between the empirical mean for the location $(1,1)$ from 5000 stochastic simulation runs and the iterative and direct coupling approximations constructed from the hybrid fluid approximation of the running example.

Based on Equation 13 we can find the cumulative rate corresponding to the mode switch to state 1 — or equivalently cumulative rate out of state 0. This is given by

$$\Lambda_1(t) = \int_0^t \mathbb{E}_{\hat{\boldsymbol{x}}|\hat{Z}^0_{s^-}}\left[q^{\boldsymbol{x}}_{\hat{Z}}(0,1)\right] ds$$

From that the cumulative distribution function corresponding to the first hitting time of state 1 of $\hat{Z}$ is given by

$$p(h_1 \leq t) = 1 - e^{-\Lambda_1(t)}$$

This is enough to the construct an approximation of the marginal behaviour of the population process $\hat{\boldsymbol{x}}$ conditioned on $\hat{Z}^1_{t^-}$. That is, for the system that does not leave the mode-switching states $\{0, 1\}$ we get

$$\begin{aligned}\frac{d}{dt}\hat{\boldsymbol{x}}(t) &= p(z = 0; t)\boldsymbol{F}^0(\hat{\boldsymbol{x}}(t)) + p(z = 1; t)\boldsymbol{F}^1(\hat{\boldsymbol{x}}(t)) \\ &= p(h_1 > t)\boldsymbol{F}^0(\hat{\boldsymbol{x}}(t)) + p(h_1 \leq t)\boldsymbol{F}^1(\hat{\boldsymbol{x}}(t))\end{aligned}$$

The general construction up to the k-th mode-switch for populations where the mode-switching process has the pure birth structure can then be given by

$$\frac{d}{dt}\hat{\boldsymbol{x}}(t) = \sum_{i=1}^{k} p(h_i \leq t)\boldsymbol{F}^i(\hat{\boldsymbol{x}}(t))$$

where the distribution $p(h_k \leq t)$ is found by constructing the mode-switching process up to the $(k-1)$-th mode giving rise to the iterative construction.

$$\begin{aligned}\Lambda_k(t) &= \int_0^t \mathbb{E}_{\hat{\boldsymbol{x}}|\hat{Z}^{k-1}_{s^-}}\left[q^{\boldsymbol{x}}_{\hat{Z}}(k-1, k)\right] ds \\ p(h_k \leq t) &= 1 - e^{-p(h_{k-1} \leq t)\Lambda_k(t)}\end{aligned} \tag{14}$$

The scheme is iterated once for all mode-switching events that can happen in a chosen finite time-interval for which the dynamics are considered. In the case of the running example with two modes and unidirectional switching this means two iterations. Alternatively, the iterative scheme can be terminated early if the probability of the next jump being considered is small in the finite time-interval.

Example 12. *Based on the discussion in this section and the trajectories from example 11 we can calculate the distribution $p(h_1 \leq t)$ and solve the system*

$$\frac{d}{dt}\hat{\boldsymbol{x}}(t) = p(h_1 > t)\boldsymbol{F}^0(\hat{\boldsymbol{x}}(t)) + p(h_1 \leq t)\boldsymbol{F}^1(\hat{\boldsymbol{x}}(t))$$

where $\boldsymbol{F}^0$ corresponds to the drift giving the random walk over the grid structure in the running example and $\boldsymbol{F}^1$ corresponds to the drift towards the target. In Figure 8 we give a comparison of the resulting probability distributions for $\hat{Z}(t) = 1$ and the relative errors for the population measure of location $(1, 1)$ between the stochastic simulation and the fluid approximation based iterative construction. The distribution for $\hat{Z}(t) = 1$ and relative errors from the direct coupling method in the previous section are given for comparison. Note that for this simple example this approach gives a more accurate representation of the distribution $\hat{Z}(t) = 1$. In addition the relative error, if discarding the values near $t = 0.0$ due to numerical inaccuracies when population levels are low, is improved. In particular, the maximum relative error is diminished and the error reaches values near zero faster. Both of the relative error trajectories feature points where the error becomes zero. These points occur when the constructed solutions cross the empirical estimate.

6 RESULTS

In this section we conduct an empirical analysis of presented direct coupling-based and iterative constructions of mean dynamics. For evaluation the approximations are compared against the empirical mean from 5000 stochastic simulations. To demonstrate the extensibility and scalability of the presented modelling and analysis ideas, we first introduce a larger model inspired by maze navigation.

EXAMPLE 13. *This example considers a 4-by-4 grid with connections between nodes constructed as shown in Figure 9. The mode transitions happen, as before, through instantaneous broadcast communication when a robot navigating the structure reaches* $(2,2)$ *or* $(3,3)$ *and detects them as targets with rate* r_s. *We assume that the robot can give the rest of the swarm enough information to reach its location. This splits the dynamics of the collective into three modes — random walk and directed walks towards locations* $(2,2)$ *and* $(3,3)$, *respectively. Approximations to the mean dynamics are then constructed based on the previous section. Analogously to the running example the entire swarm is assumed to start at location* $(0,0)$.

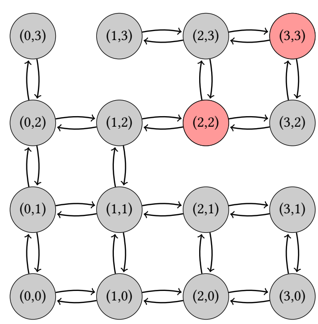


Fig. 9. Spatial structure of the maze example.

The maze navigation example presented above has an absorbing state corresponding to all the robots being in location $(3,3)$. In contrast, the following provides an example where there is no such absorbing state.

EXAMPLE 14. *The second example we are going to consider extends the running example, set up in Example 5, with an additional mode switch. Recall that the initial dynamic mode in this example is given by a random walk on the graph structure over a 2-by-2 grid in Figure 3. We are going to consider the case where, from the second mode describing a directed walk towards* $(1,1)$, *the population dynamics can revert back to the initial mode. After reverting to initial random walk dynamics the dynamics will no longer change. We are going to say that the change from the second mode, defined by the directed walk towards* $(1,1)$, *to the third mode, defined by the random walk, is caused again by a broadcast action sent out by a robot at location* $(1,1)$ *happening at rate* $0.1r_s$.

Let us consider the direct coupling and iterative methods for the two examples at 5 different population levels (100, 200, 300, 500, 1000). In order to study the behaviour of the approximations under different parametrisations we consider, for each population level, 50 parameter values for r_m and r_s that are randomly sampled from the interval

$(0, 1)$. Thus we consider 50 different models operating over the same graph structure, and for each of these we consider five different population levels, giving 250 models in total. The resulting approximations are then compared to mean population measures acquired from 5000 stochastic simulation runs.

Table 1 gives an overview of the mean computation times for Example 13 based on the maze navigation example. The experiments were implemented using DifferentialEquations.jl package [31] for the Julia programming language and run in batches corresponding to the population size. In each case the timings of the calculations for first parametrisations were not considered as they include the just-in-time compilation overhead of Julia.

Table 1. Comparison of mean computation times for the maze example (seconds).

Population size N	Direct coupling	Iterative	Sampled trajectories (5000)
100	0.0010	0.15	2.01
200	0.0011	0.14	7.12
300	0.0013	0.15	10.95
500	0.0011	0.15	16.56
1000	0.0011	0.15	27.22

Table 2 presents an error analysis for the two examples considered in this section. For the maze example we consider the approximations to the expected population at location $(3, 3)$ while for the 2-by-2 example we consider the location $(1, 1)$. For each experiment corresponding to a parametrisation of r_m, r_s and choice of population level N we calculate the maximum absolute errors with respect to the empirical mean derived from corresponding stochastic simulation runs. In order to make the absolute errors comparable we have normalised them based on the population size to reflect the error in terms of the proportion of considered population. For each of the population levels the table displays the mean and standard deviation of these maximum errors as well as the largest of the calculated maximum errors. As we can see from Table 1 the iterative method creates some computational overhead in the implementation. At the same time, based on Table 2, in most of the tested cases the iterative construction offers in most case better accuracy than directly coupling the approximate filtering equations with the fluid drifts. Finally, Figures 10 and 11 give the corresponding error surfaces for the 4-by-4 maze and the three mode 2-by-2 example respectively at population levels 100 and 200. In particular, maximum errors for each of the parametrisations of r_m and r_s are plotted. These indicate that the iterative construction deals better with the parametrisations where the rate of movement parameter r_m is much lower than the parameter r_s defining the rate of broadcasting. The precise reasons why the iterative construction gives an improved estimate for the mean are proposed as a subject for further research.

7 RELATED WORK

The construction proposed in this paper will result in a specific class of models which we call mode-switching population systems. These systems have many similarities with bistable or hybrid behaviour in the context of models of biochemical reactions

Table 2. Comparison of mean maximum errors. For the maze example the errors are calculated for location $(3, 3)$. For the 2-by-2 example with three modes the errors are calculated for location $(1, 1)$.

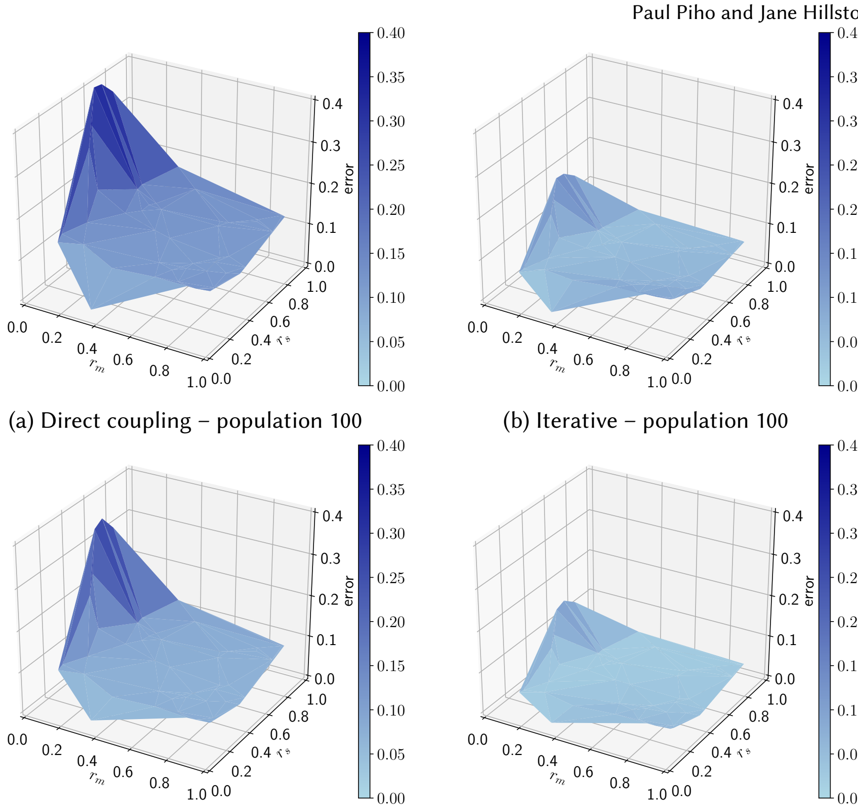
		4-by-4 maze mean (sd) max.	max.	2-by-2 three modes mean (sd) max.	max.
		$t \in [0, 100.0]$		$t \in [0, 100.0]$	
$N = 100$	Direct coupling	14.5 (7.2)%	39.4%	2.4 (2.9)%	13.5%
	Iterative	7.3 (3.2)%	17.4%	1.7 (1.6)%	6.9%
$N = 200$	Direct coupling	9.9 (6.3)%	33.7%	1.5 (1.8)%	8.5%
	Iterative	4.3 (2.5)%	13.5%	1.2 (1.3)%	5.8%
$N = 300$	Direct coupling	7.8 (5.2)%	28.7%	1.1 (1.5)%	6.9%
	Iterative	3.2 (1.9)%	10.1%	1.0 (1.1)%	5.1%
		$t \in [0, 100.0]$		$t \in [0, 10.0]$	
$N = 500$	Direct coupling	5.9 (4.2)%	21.9%	0.8 (1.0)%	4.7%
	Iterative	2.2 (1.4)%	7.1%	0.7 (1.0)%	4.1%
$N = 1000$	Direct coupling	4.0 (3.0)%	15.7%	0.6 (0.7)%	3.3%
	Iterative	1.4 (1.0)%	4.9%	0.5 (0.8)%	3.4%

under the presence of low-copy number components. The terminology used in this paper is chosen to emphasise the fact that the way in which such dynamics arise from the population processes differs. In particular, we are not considering the usual constructions where the state space of the pCTMC is partitioned into parts describing the evolution of high and low-copy number components (usually referred to as species in biochemical literature). Nevertheless, the approximate models that arise from our constructions have obvious counterparts and existing analysis methods. These are commonly referred to by the name of hybrid methods and it is prudent to give a short overview of a few of them.

The hybrid methods can be categorised based on how the variables describing the evolution of high-copy number components are treated. The usual approach is that the variables corresponding to high-copy components get a continuous description. For example, [19, 25] derive expressions that represent the mean behaviour of abundances of high-copy components. In the case of [19] models considered assume that the discrete stochastic variables are mutually independent and independent of continuous variables. The conditional mean equations derived in [25] on the other hand coincide with the method of conditional moments presented in [18] and give rise to a system of differential algebraic equations. Solutions to the resulting systems of differential algebraic equations are generally not trivial. Finally, the papers [11, 33] consider the linear noise approximation coupled with discrete stochastic evolution of the low-copy components and make use of the time-scale separation to give approximate closed-form solutions to arising hybrid models.

8 CONCLUSION

Fluid approximations are a powerful technique for scalable treatment of population CTMC dynamics. However, existing results rely on assumptions which are not shared by all population systems. For example, broadcast communication and information



(a) Direct coupling – population 100

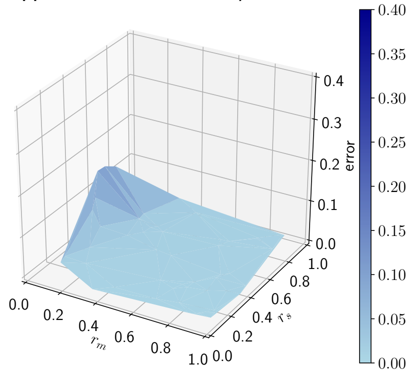
(b) Iterative – population 100

(c) Direct coupling – population 200

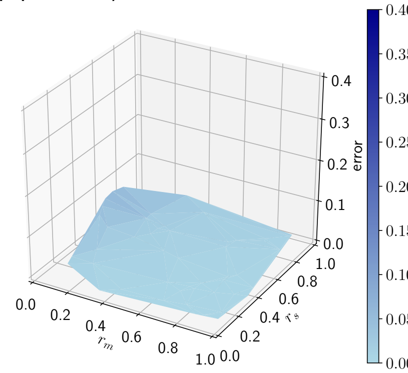
(d) Iterative – population 200

Fig. 10. Error surfaces for the maze example. Displays the randomly sampled pairs of r_m and $_s$ against the corresponding maximum errors for location $(3, 3)$.

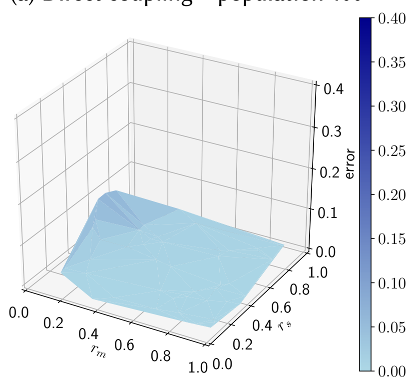
cascades provide an interesting modelling challenge where actions of single agents can cause changes in the macro-level dynamics of the population. This is motivated by scenarios where the population accumulates knowledge about its operating environment throughout the course of its evolution. Mode-switching dynamics relating to information or knowledge spread include subtle but important differences from the more commonly studied bi-stable behaviour of pCMTCs. In the context of broadcast communication we have presented a framework in which to study CTMC models of population dynamics where the collective behaviour of the population goes through mode and demonstrated how fluid approximation based analysis can be adapted to such scenarios. Finally, we have introduced two constructions for approximate marginal population dynamics based on the hybrid fluid approximation results and presented a comparison between them. These show that methods that make explicit use of special structures of the mode-switching process, like the iterative construction of mean dynamics, may in some cases offer improved accuracy and should be of interest for further study.



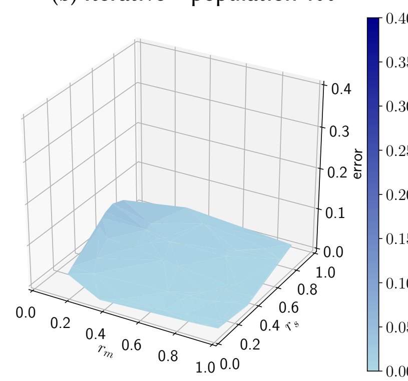
(a) Direct coupling – population 100



(b) Iterative – population 100



(c) Direct coupling – population 200



(d) Iterative – population 200

Fig. 11. Error surfaces for the three mode 2-by-2 grid example. Displays the randomly sampled pairs of r_m and $_s$ against the corresponding maximum errors for the location $(1, 1)$.

The methods presented open up fluid approximation results for model checking and control in cases where density-dependence is violated relatively infrequently during the transient evolution. However, there are several interesting questions that remain open for further investigation. For example, what are the precise conditions under which the presented iterative construction can be expected to offer an improved approximation. Additionally, as further work we aim to study the behaviour and limitations in situations where the mode switching process can exhibit branching, for example, based on the uncertain order in which the knowledge is acquired. Finally, we are considering applications of central-limit approximation [16] and moment-closure approximations [24] to gain a better description of the stochastic behaviour of the transient evolution of population systems.

Acknowledgements

This work was supported by EPSRC grant EP/L01503X/1 (CDT in Pervasive Parallelism).

REFERENCES

[1] Behçet Açikmese and David S. Bayard. 2012. A Markov chain approach to probabilistic swarm guidance. In *American Control Conference, ACC 2012, Montreal, QC, Canada, June 27-29, 2012*. IEEE, 6300–6307. http://ieeexplore.ieee.org/document/6314729/

[2] Marco Ajmone Marsan, Gianni Conte, and Gianfranco Balbo. 1984. A Class of Generalized Stochastic Petri Nets for the Performance Evaluation of Multiprocessor Systems. *ACM Trans. Comput. Syst.* 2, 2 (May 1984), 93–122. https://doi.org/10.1145/190.191

[3] Julien Bect. 2010. A unifying formulation of the Fokker–Planck–Kolmogorov equation for general stochastic hybrid systems. *Nonlinear Analysis: Hybrid Systems* 4, 2 (2010), 357 – 370. https://doi.org/10.1016/j.nahs.2009.07.008 IFAC World Congress 2008.

[4] Marco Bernardo and Roberto Gorrieri. 1996. Extended Markovian Process Algebra. In *CONCUR '96: Concurrency Theory*, Ugo Montanari and Vladimiro Sassone (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 315–330.

[5] Bortolussi and Jane Hillston. 2015. Model checking single agent behaviours by fluid approximation. *Inf. Comput.* 242 (2015), 183–226. https://doi.org/10.1016/j.ic.2015.03.002

[6] Luca Bortolussi. 2010. Limit Behavior of the Hybrid Approximation of Stochastic Process Algebras. In *Analytical and Stochastic Modeling Techniques and Applications, 17th International Conference, ASMTA 2010, Cardiff, UK, June 14-16, 2010. Proceedings*. 367–381. https://doi.org/10.1007/978-3-642-13568-2_26

[7] Luca Bortolussi. 2016. Hybrid behaviour of Markov population models. *Inf. Comput.* 247 (2016), 37–86. https://doi.org/10.1016/j.ic.2015.12.001

[8] Luca Bortolussi, Luca Cardelli, Marta Kwiatkowska, and Luca Laurenti. 2019. Central Limit Model Checking. *ACM Trans. Comput. Log.* 20, 4 (2019), 19:1–19:35. https://doi.org/10.1145/3331452

[9] Luca Bortolussi and Roberta Lanciani. 2014. Stochastic Approximation of Global Reachability Probabilities of Markov Population Models. In *Computer Performance Engineering - 11th European Workshop, EPEW 2014, Florence, Italy, September 11-12, 2014. Proceedings*. 224–239. https://doi.org/10.1007/978-3-319-10885-8_16

[10] Leo Bronstein and Heinz Koeppl. 2018. Marginal process framework: A model reduction tool for Markov jump processes. *Phys. Rev. E* 97 (Jun 2018), 062147. Issue 6. https://doi.org/10.1103/PhysRevE.97.062147

[11] Luca Cardelli, Marta Kwiatkowska, and Luca Laurenti. 2016. A Stochastic Hybrid Approximation for Chemical Kinetics Based on the Linear Noise Approximation. In *Computational Methods in Systems Biology - 14th International Conference, CMSB 2016, Cambridge, UK, September 21-23, 2016, Proceedings*. 147–167. https://doi.org/10.1007/978-3-319-45177-0_10

[12] A. Crudu, A. Debussche, A. Muller, and O. Radulescu. 2012. convergence Of Stochastic Gene Networks To Hybrid Piecewise Deterministic Processes. *The Annals of Applied Probability* 22, 5 (2012), 1822–1859.

[13] M.H.A. Davis. 1993. *Markov Models & Optimization.* Taylor & Francis.

[14] Vaibhav Deshmukh, Karthik Elamvazhuthi, Shiba Biswal, Zahi Kakish, and Spring Berman. 2018. Mean-Field Stabilization of Markov Chain Models for Robotic Swarms: Computational Approaches and Experimental Results. *IEEE Robotics and Automation Letters* 3, 3 (2018), 1985–1992. https://doi.org/10.1109/LRA.2018.2792696

[15] Karthik Elamvazhuthi and Spring Berman. 2019. Mean-field models in swarm robotics: a survey. *Bioinspiration & Biomimetics* 15, 1 (nov 2019), 015001. https://doi.org/10.1088/1748-3190/ab49a4

[16] Stewart N. Ethier and Thomas G. Kurtz. 1986. *Markov processes – characterization and convergence.* John Wiley & Sons Inc., New York. 534 pages.

[17] Daniel T. Gillespie. 1977. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* 81, 25 (Dec. 1977), 2340–2361. https://doi.org/10.1021/j100540a008

[18] J. Hasenauer, V. Wolf, A. Kazeroonian, and F. J. Theis. 2014. Method of conditional moments (MCM) for the Chemical Master Equation. *Journal of Mathematical Biology* 69, 3 (01 Sep 2014), 687–735. https://doi.org/10.1007/s00285-013-0711-5

[19] Andreas Hellander and Per Lötstedt. 2007. Hybrid method for the chemical master equation. *J. Comput. Physics* 227, 1 (2007), 100–122. https://doi.org/10.1016/j.jcp.2007.07.020

[20] F. Horn and R. Jackson. 1972. General mass action kinetics. *Archive for Rational Mechanics and Analysis* 47, 2 (01 Jan 1972), 81–116. https://doi.org/10.1007/BF00251225

[21] Thomas G. Kurtz. 1970. Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes. *Journal of Applied Probability* 7, 1 (1970), 49–58.

[22] Thomas G. Kurtz. 1972. The Relationship between Stochastic and Deterministic Models for Chemical Reactions. *The Journal of Chemical Physics* 57, 7 (1972), 2976–2978. https://doi.org/10.1063/1.1678692
[23] Michele Loreti and Jane Hillston. 2016. Modelling and Analysis of Collective Adaptive Systems with CARMA and its Tools. In *Formal Methods for the Quantitative Evaluation of Collective Adaptive Systems, Advanced Lectures*. 83–119. https://doi.org/10.1007/978-3-319-34096-8_4
[24] Donald A. McQuarrie. 1963. Kinetics of Small Systems. I. *The Journal of Chemical Physics* 38, 2 (1963), 433–436. https://doi.org/10.1063/1.1733676
[25] Stephan Menz, Juan C. Latorre, Christof Schütte, and Wilhelm Huisinga. 2012. Hybrid Stochastic-Deterministic Solution of the Chemical Master Equation. *Multiscale Modeling & Simulation* 10, 4 (2012), 1232–1262. https://doi.org/10.1137/110825716
[26] Michalis Michaelides, Jane Hillston, and Guido Sanguinetti. 2019. Geometric fluid approximation for general continuous-time Markov chains. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 475, 2229 (2019). https://doi.org/10.1098/rspa.2019.0100
[27] M. K. Molloy. 1982. Performance Analysis Using Stochastic Petri Nets. *IEEE Trans. Comput.* 31, 9 (Sept. 1982), 913–917. https://doi.org/10.1109/TC.1982.1676110
[28] J. R. Norris. 1998. *Markov chains*. Cambridge University Press, Cambridge.
[29] Bernt Oksendal. 1992. *Stochastic Differential Equations (3rd Ed.): An Introduction with Applications*. Springer-Verlag, Berlin, Heidelberg.
[30] Paul Piho and Jane Hillston. 2018. Policy Synthesis for Collective Dynamics. In *Quantitative Evaluation of Systems - 15th International Conference, QEST 2018, Beijing, China, September 4-7, 2018, Proceedings*. 356–372. https://doi.org/10.1007/978-3-319-99154-2_22
[31] Chris Rackauckas and Qing Nie. 2017. DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *Journal of Open Research Software* 5 (05 2017). https://doi.org/10.5334/jors.151
[32] David Schnoerr, Guido Sanguinetti, and Ramon Grima. 2017. Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *Journal of Physics A: Mathematical and Theoretical* 50, 9 (jan 2017), 093001. https://doi.org/10.1088/1751-8121/aa54d9
[33] Philipp Thomas, Nikola Popović, and Ramon Grima. 2014. Phenotypic switching in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America* 111 (04 2014). https://doi.org/10.1073/pnas.1400049111

A FILTERING DISTRIBUTION

We are going to work with the hybrid fluid limit PDMP $(\hat{\boldsymbol{x}}, \hat{Z})$, introduced in Section 4, of a mode-switching population system and derive the evolution equation for the filtering distribution $\pi_t(z')$

$$\frac{d}{dt}\pi_t(z') = \frac{d}{dt}p(z'; t \mid \boldsymbol{x}_{t^-}) \tag{15}$$

corresponding to the probability that the mode-switching process $\hat{Z}$ is in state z' given a sample trajectory $\boldsymbol{x}_{t^-}$ of the population process. As mentioned in the main text of this paper the derivation of the filtering distribution follows the same outline as given for continuous time Markov chains in [10] and is presented here mainly for completeness of presentation. The piece of background information needed in the following is how the transition densities $p(\hat{\boldsymbol{x}}, \hat{z}; t \mid \boldsymbol{x}, z; s)$ of the PDMPs considered in this paper depend on time t. This is characterised by the following forward equation derived for example in [3].

$$
\begin{aligned}
\frac{\partial}{\partial t} p(\hat{\boldsymbol{x}}, \hat{z}; t \mid \boldsymbol{x}, z; s) &= \mathcal{L}^* p(\hat{\boldsymbol{x}}, \hat{z}; t \mid \boldsymbol{x}, z; s) \\
&= -\sum_i \partial_i \boldsymbol{F}_i^z(\boldsymbol{x}) p(\boldsymbol{x}, z'; t \mid \boldsymbol{x}_s, z; s) \\
&\quad + \sum_{k, k \neq \hat{z}} \Big[p(\hat{\boldsymbol{x}}, k; t \mid \boldsymbol{x}, z; s) q_Z^{\boldsymbol{x}}(k, \hat{z}) - p(\hat{\boldsymbol{x}}, k; t \mid \boldsymbol{x}, z; s) q_Z^{\boldsymbol{x}}(\hat{z}, k) \Big]
\end{aligned}
$$

Evaluated at $t = s$ the above gives us

$$
\mathcal{L}^* p(\hat{\boldsymbol{x}}, \hat{z}; s \mid \boldsymbol{x}, z; s) = \begin{cases} -\sum_i \partial_i \boldsymbol{F}_i^z(\boldsymbol{x}) + \sum_{k, k \neq z} \left[q_Z^{\boldsymbol{x}}(k, z) - q_Z^{\boldsymbol{x}}(z, k) \right] & \text{for } \hat{\boldsymbol{x}} = \boldsymbol{x}, \hat{z} = z \\ 0 & \text{otherwise} \end{cases}
$$

The differences between the derivation in [10], that makes our case somewhat simpler, is that we are only going to consider continuous sample paths $\boldsymbol{x}_{t^-}$. The reason for that is given in Section 4 where we assumed that the transitions affecting the mode-switching process do no change the state of the population process directly. In order to derive the Equation 15 we are going to consider the value of $\pi_{t+\Delta t}(z')$ and directly apply the definition of a derivative. First of all, by leveraging the Bayes' rule, we get

$$
\begin{aligned}
\pi_{t+\Delta t}(z') &= p(z'; t + \Delta t \mid \boldsymbol{x}_{(t+\Delta t)^-}) \\
&= \sum_z p((z'; t + \Delta t) \cap (z; t) \mid \boldsymbol{x}_{t^-} \cap \boldsymbol{x}_{t+\Delta t}) \\
&= \sum_z \frac{p(z', \boldsymbol{x}_{t+\Delta t}; t + \Delta t \mid z, \boldsymbol{x}_{t^-}; t) \pi_t(z)}{p(\boldsymbol{x}_{t+\Delta t}; t + \Delta t \mid \boldsymbol{x}_{t^-})} \\
&= \sum_z \frac{\pi_t(z) \left[p(z', \boldsymbol{x}_t; t \mid z, \boldsymbol{x}_t; t) + \Delta t \mathcal{L}^* p(z', \boldsymbol{x}_t; t \mid z, \boldsymbol{x}_t; t) + o(\Delta t) \right]}{1 + \Delta t \mathbb{E}_{\hat{Z} \mid \boldsymbol{x}_{t^-}} \left[-\sum_i \partial_i \boldsymbol{F}_i^{\hat{Z}}(\boldsymbol{x}) \right] + o(\Delta t)}
\end{aligned}
$$

The last line results from the Taylor expansion of $p(z', \boldsymbol{x}_{t+\Delta t}; t + \Delta t \mid z, \boldsymbol{x}_{t^-}; t)$ around t. Putting the above into the definition of a derivative we get

$$
\begin{aligned}
\frac{d}{dt} \pi_t(z') &= \lim_{\Delta t \to 0} \frac{\pi_{t+\Delta t}(z') - \pi_t(z')}{\Delta t} = \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Bigg[\sum_z \frac{\pi_t(z) \left[p(z', \boldsymbol{x}_t; t \mid z, \boldsymbol{x}_{t^-}; t) + \Delta t \mathcal{L}^* p(z', \boldsymbol{x}_t; t \mid z, \boldsymbol{x}_t; t) + o(\Delta t) \right]}{1 + \Delta t \mathbb{E}_{\hat{Z} \mid \boldsymbol{x}_{t^-}} \left[-\sum_i \partial_i \boldsymbol{F}_i^{\hat{Z}}(\boldsymbol{x}) \right] + o(\Delta t)} \\
&\quad - \frac{\pi_t(z') \left[1 + \Delta t \mathbb{E}_{\hat{Z} \mid \boldsymbol{x}_{t^-}} \left[-\sum_i \partial_i \boldsymbol{F}_i^{\hat{Z}}(\boldsymbol{x}) \right] + o(\Delta t) \right]}{1 + \Delta t \mathbb{E}_{\hat{Z} \mid \boldsymbol{x}_{t^-}} \left[-\sum_i \partial_i \boldsymbol{F}_i^{\hat{Z}}(\boldsymbol{x}) \right] + o(\Delta t)} \Bigg]
\end{aligned} \tag{16}
$$

Now let us concentrate on the numerator of the fraction inside the brackets. First of all note that $p(z', \boldsymbol{x}_t; t \mid z, \boldsymbol{x}_{t^-}; t)$ is 1 exactly when $z' = z$ and zero otherwise. In particular, we get

$$\begin{aligned}\pi_t(z') + \sum_z \pi_t(z)\ [\Delta t \mathcal{L}^* p(z', \boldsymbol{x}_t; t \mid z, \boldsymbol{x}_t; t) + o(\Delta t)]\\ - \pi_t(z')\Big[1 + \Delta t \mathbb{E}_{\hat{Z}|\boldsymbol{x}_{t^-}}\Big[-\sum_i \partial_i \boldsymbol{F}_i^{\hat{Z}}(\boldsymbol{x})\Big] + o(\Delta t)\Big]\\ = \sum_z \pi_t(z)\ [\Delta t \mathcal{L}^* p(z', \boldsymbol{x}_t; t \mid z, \boldsymbol{x}_t; t) + o(\Delta t)]\\ - \pi_t(z')\Big[\Delta t \mathbb{E}_{\hat{Z}|\boldsymbol{x}_{t^-}}\Big[-\sum_i \partial_i \boldsymbol{F}_i^{\hat{Z}}(\boldsymbol{x})\Big] + o(\Delta t)\Big]\end{aligned}$$

The above gives the numerator in Equation 16. Multiplying the numerator by $\frac{1}{\Delta t}$ and taking the limit $\Delta t \to 0$ in both the numerator and denominator of Equation 16 then gives us

$$\begin{aligned}\frac{d}{dt}\pi_t(z') &= \sum_z \pi_t(z)\mathcal{L}^* p(z', \boldsymbol{x}_t; t \mid z, \boldsymbol{x}_t; t) - \pi_t(z')\mathbb{E}_{\hat{Z}|\boldsymbol{x}_{t^-}}\Big[\sum_i \partial_i \boldsymbol{F}_i^{\hat{Z}}(\boldsymbol{x})\Big]\\ &= \sum_z \pi_t(z) q(z, z') - \pi_t(z')\Big[\sum_i \partial_i \boldsymbol{F}_i^{\hat{z}}(\boldsymbol{x}) - \mathbb{E}_{\hat{Z}|\boldsymbol{x}_{t^-}}\Big[\sum_i \partial_i \boldsymbol{F}_i^{\hat{Z}}(\boldsymbol{x})\Big]\Big]\end{aligned}$$

as claimed in Section 5.